



OIPE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/655,160

DATE: 01/16/2003

TIME: 13:57:46

Input Set : A:\PTO.VSK.txt

Output Set: N:\CRF4\01162003\I655160.raw

2 <110> APPLICANT: CALIFORNIA INSTITUTE OF TECHNOLOGY
 3 Paul, STERNBERG W.
 4 Maureen, EARR M.
 5 <130> TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE HOMOLOGS REQUIRED FOR MALE
 MATING BEHAVIOR

6 IN NEMATODES AND ASSAYS BASED THEREON
 7 <130> FILE REFERENCE: CIT1520-2

11 <140> CURRENT APPLICATION NUMBER: US 09/655,160

C--> 12 <141> CURRENT FILING DATE: 2002-12-30

14 <150> PRIORITY APPLICATION NUMBER: US 09/479,467

15 <151> PRIORITY FILING DATE: 2000-01-06

17 <150> PRIORITY APPLICATION NUMBER: US 60/115,127

18 <151> PRIORITY FILING DATE: 1999-01-06

20 <160> NUMBER OF SEQ ID NOS: 16

22 <170> SOFTWARE: PatentIn version 3.1

24 <210> SEQ ID NO: 1

25 <211> LENGTH: 12012

26 <212> TYPE: DNA

27 <213> ORGANISM: Homo sapiens PKD-1 gene

28 <220> FEATURE:

30 <221> NAME/KEY: CDS

31 <222> LOCATION: (1)..(12912)

32 <223> OTHER INFORMATION:

W--> 34 <400> 1

```

35 atg tgg ccc gcc agc ccc gcc cgc ctg gcc ctg gcc ctg gcc 48
36 Met Pro Pro Ala Ala Pro Ala Arg Leu Ala Leu Ala Leu Gly Leu Gly
37 1 10 15
38 ctg tgg ctc gcc agc ctg gcc gcc ggc ccc ggc gcc tgc ggc ccc 96
39 Ser Trp Leu Gly Ala Leu Ala Gly Gly Pro Gly Arg Gly Cys Gly Pro
40 20 25 30
41 tgc gag ccc ccc tgc ctc tgc gcc cca gcc ccc gcc gcc tgc gcc 144
42 Cys Glu Pro Pro Cys Leu Cys Gly Pro Ala Pro Gly Ala Ala Cys Arg
43 35 40 45
44 gtc aac tgc tgg gcc cgc gcc ctg gcc aac ctc ggt ccc gcc ctg gcc 192
45 Val Asn Cys Ser Gly Arg Gly Leu Arg Thr Leu Gly Pro Ala Leu Arg
46 50 55 60
47 atc ccc gcc gcc gcc aca gag cta gcc gtc tcc ccc aac ctg ctc cgc 240
48 Ile Pro Ala Asp Ala Thr Glu Leu Asp Val Ser His Asn Leu Leu Arg
49 65 70 75 80
50 gcc ctg gcc att gcc ctc ctg gcc aac ctc tgg gcc ctg gcc gag ctg 288
51 Ala Leu Asp Val Gly Leu Leu Ala Asn Leu Ser Ala Leu Ala Glu Leu
52 85 90 95
53 gat ata gcc aac aac aag att tct aag tta gaa gaa gga ata ttt gct 336
54 Asp Ile Ser Asn Asn Lys Ile Ser Thr Leu Glu Glu Gly Ile Phe Ala

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/655,160

DATE: 01/16/2003

TIME: 13:57:46

Input Set : A:\PTO.VSK.txt

Output Set : N:\CRF4\01162003\I655160.raw

61	100	105	110	
63	aat tta ttt aat tta agt gaa ata aac ctg agt ggg aac ccg ttt gag	384		
64	Asn Leu Phe Asn Leu Ser Glu Ile Asn Leu Ser Gly Asn Pro Phe Glu			
65	115	120	125	
67	tgt gac tgt gcc ctg gcc tgg ctg ccg caa tgg gcc gag gag cag cag	432		
68	Cys Asp Cys Gly Leu Ala Trp Leu Pro Gln Trp Ala Glu Glu Gln G n			
69	130	135	140	
71	gtg ccg gtg gtg cag ccc gag gca gcc acg tgt gct ggg cct gcc tcc	480		
72	Val Arg Val Val Gln Pro Glu Ala Ala Thr Cys Ala Gly Pro Gly Ser			
73	145	150	155	160
75	cct gct ccg cag cct ctg ctt gcc atc ccc tgg ctg gac agt ggt tgt	528		
76	Leu Ala Gly Gln Pro Leu Leu Gly Ile Pro Leu Leu Asp Ser Gly Cys			
77	165	170	175	
79	ggt gag gac tat gtc gcc tgc ctg cct gac aac agc tca gcc acc gtg	576		
80	Gly Glu Glu Tyr Val Ala Cys Leu Pro Asp Asn Ser Ser Gly Thr Val			
81	180	185	190	
83	gca gca gtg tcc ttt tca gct gcc cag gaa gcc ctg ctt cag cca gag	624		
84	Ala Ala Val Ser Phe Ser Ala Ala His Glu Gly Leu Leu Gln Pro Glu			
85	195	200	205	
87	gcc tgc agc gcc ttc tcc ttc tcc acc gcc cag gcc ctg gca gcc ctg	672		
88	Ala Cys Ser Ala Phe Cys Phe Ser Thr Gly Gln Gly Leu Ala Ala Leu			
89	210	215	220	
91	tgg gag cag gcc tgg tcc ctg tgt ggg gcc gcc cag ccc tcc agt gcc	720		
92	Ser Glu Gln Gly Trp Cys Leu Cys Gly Ala Ala Gln Pro Ser Ser Ala			
93	225	230	235	240
95	tcc ttt gcc tgc ctg tcc ctg tcc tcc ggg ccc ccg gca cct cct gcc	768		
96	Ser Phe Ala Cys Leu Ser Leu Cys Ser Gly Pro Pro Ala Pro Pro Ala			
97	245	250	255	
99	ccc acc tgt agg gcc ccc acc ctg ctg cag cag gtc ttc cct gcc tcc	816		
100	Pro Thr Cys Arg Gly Pro Thr Leu Leu Gln His Val Phe Pro Ala Ser			
101	260	265	270	
103	cca ggg gcc acc ctg tgg ggg ccc cag gga cct ctg gcc tct gcc cag	864		
104	Pro Gly Ala Thr Leu Val Gly Pro His Gly Pro Leu Ala Ser Gly Gln			
105	275	280	285	
107	cta gca gcc ttc cag atc gct gcc ccg ctg cct gtc act gac aca gcc	912		
108	Leu Ala Ala Phe His Ile Ala Ala Pro Leu Pro Val Thr Asp Thr Arg			
109	290	295	300	
111	tgg gac ttc gga gac gcc tcc gcc gag gtg gat gcc gct ggg ccg gct	960		
112	Trp Asp Phe Gly Asp Gly Ser Ala Glu Val Asp Ala Ala Gly Pro Ala			
113	305	310	315	320
115	gcc tgg cat gcc tat tgg ctg cct ggg ccg tat cag gtg acc gcc gtg	1008		
116	Ala Ser His Arg Tyr Val Leu Pro Gly Arg Tyr His Val Thr Ala Val			
117	325	330	335	
119	ctg gcc ctg ggg gcc ggc tca gcc ctg ctg ggg aca gac gtg cag gtg	1056		
120	Leu Ala Leu Gly Ala Gly Ser Ala Leu Leu Gly Thr Asp Val Gln Val			
121	340	345	350	
123	gaa gcc gca cct gcc gcc ctg gag ctg ctg tgc ccg tcc tcc gtg cag	1104		
124	Glu Ala Ala Pro Ala Ala Leu Glu Leu Val Cys Pro Ser Ser Val Gln			
125	355	360	365	

RAW SEQUENCE LISTING

DATE: 01/16/2003

PATENT APPLICATION: US/09/655,160

TIME: 13:57:46

Input Set : A:\PTO.VSK.txt

Output Set: N:\CRF4\01162003\I655160.raw

127	agt	gac	gag	agg	ctc	gac	ctc	agg	atc	cag	aac	ggc	ggt	ggc	tca	ggc	1152
128	Ser	Asp	Glu	Ser	Leu	Asp	Leu	Ser	Ile	Gln	Asn	Arg	Gly	Gly	Ser	Gly	
129		375					375					380					
131	ctg	gag	gcc	gac	tac	agg	atc	ctg	ggc	ctg	ggc	gag	gag	ccg	gcc	cca	1200
132	Leu	Glu	Ala	Ala	Tyr	Ser	Ile	Val	Ala	Leu	Gly	Glu	Glu	Pro	Ala	Arg	
133	385					390					395				400		
135	ggg	gtg	cac	ccg	ctc	tgc	ccc	tcc	gac	acc	gac	atc	ttc	ccc	ggc	aac	1248
136	Ala	Val	His	Pro	Leu	Cys	Pro	Ser	Asp	Thr	Gln	Ile	Phe	Pro	Gly	Asn	
137				405					410						415		
139	ggg	cac	tgc	tac	ggc	ctc	gtg	ctg	gag	aag	ggc	ggc	tgg	ctc	cag	ggc	1296
140	Gly	His	Cys	Tyr	Arg	Leu	Val	Val	Glu	Lys	Ala	Ala	Trp	Leu	Gln	Ala	
141				420					425						430		
143	cag	cag	cac	tgt	cag	gcc	tgg	ccc	ggg	ggc	ggc	ctc	gca	atc	gtg	gac	1344
144	Gln	Glu	Glu	Cys	Gln	Ala	Trp	Ala	Gly	Ala	Ala	Leu	Ala	Met	Val	Asp	
145				435					440						445		
147	agt	ccc	gcc	gtg	cag	ggc	ttc	ctg	gtc	tcc	ggg	gtc	acc	agg	agg	cta	1392
148	Ser	Pro	Ala	Val	Gln	Arg	Phe	Leu	Val	Ser	Arg	Val	Thr	Arg	Ser	Leu	
149				450					455						460		
151	gac	gtg	tgc	atc	ggc	ttc	tcc	act	gtc	cag	ggg	gtg	gag	gtg	ggc	cca	1440
152	Asp	Val	Trp	Ile	Gly	Phe	Ser	Thr	Val	Gln	Gly	Val	Glu	Val	Gly	Pro	
153	465					470					475				480		
155	ggg	ccc	cag	ggg	gag	gcc	ttc	agg	ctg	gag	agg	tgc	cag	aac	tgg	ctg	1488
156	Ala	Pro	Glu	Gly	Glu	Ala	Phe	Ser	Leu	Glu	Ser	Cys	Gln	Asn	Trp	Leu	
157				485					490						495		
159	ccc	ggg	gac	ccc	cac	ccc	ggc	ccc	ggc	gag	cac	tgc	gtc	ccc	ctc	ggg	1536
160	Pro	Gly	Glu	Pro	His	Pro	Ala	Thr	Ala	Glu	His	Cys	Val	Arg	Leu	Gly	
161				500					505						510		
163	ccc	acc	ggg	tgg	tgt	cac	acc	gac	ctg	tgc	tca	ggg	ccc	cac	agg	tac	1584
164	Pro	Thr	Gly	Trp	Cys	Asn	Thr	Asp	Leu	Cys	Ser	Ala	Pro	His	Ser	Tyr	
165				515					520						525		
167	gtc	tgc	gac	ctg	cag	ccc	gga	ggc	ccc	gtg	cag	gat	ggc	gac	aac	ctc	1632
168	Val	Cys	Glu	Leu	Gln	Pro	Gly	Gly	Pro	Val	Gln	Asp	Ala	Glu	Asn	Leu	
169				530					535						540		
171	ctc	gtg	ggc	ggg	ccc	agt	ggg	gac	ctg	cag	gga	ccc	ctg	acc	ccc	ctg	1680
172	Leu	Val	Gly	Ala	Pro	Ser	Gly	Asp	Leu	Gln	Gly	Pro	Leu	Thr	Pro	Leu	
173	545					550					555				560		
175	gca	cag	cac	gac	ggc	ctc	tca	ggc	ccc	cac	gag	ccc	gtg	gag	gtc	ctg	1728
176	Ala	Gln	Glu	Asp	Gly	Leu	Ser	Ala	Pro	His	Glu	Pro	Val	Glu	Val	Met	
177				565					570						575		
179	gta	ttc	ccc	ggc	ctg	ggc	ctg	agg	gaa	ggc	ctc	ctc	acc	agg	ggc		1776
180	Val	Phe	Pro	Gly	Leu	Arg	Leu	Ser	Arg	Glu	Ala	Phe	Leu	Thr	Thr	Ala	
181				580					585						590		
183	gaa	ttt	ggc	acc	cag	gag	ctc	ggg	ggg	ccc	ggc	cag	ctg	ggc	ctg	cag	1824
184	Glu	Phe	Gly	Thr	Gln	Glu	Leu	Arg	Arg	Pro	Ala	Gln	Leu	Arg	Leu	Gln	
185				595					600						605		
187	gtg	tac	ggc	ctc	ctc	agg	aca	ggg	acc	ccc	gag	aac	ggc	agg	gag		1872
188	Val	Tyr	Arg	Leu	Leu	Ser	Thr	Ala	Gly	Thr	Pro	Glu	Asn	Gly	Ser	Glu	
189				610											615		
191	ccc	cag	agg	agg	tcc	ccc	gac	aac	agg	acc	cag	ctg	ggc	ccc	ggc	tgc	1920

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/655,160

DATE: 01/16/2003

TIME: 13:57:46

Input Set : A:\PTO.VSK.txt

Output Set: N:\CRF4\01162003\I655160.raw

```

192 Pro Glu Ser Arg Ser Pro Asp Asn Arg Thr Gln Leu Ala Pro Ala Cys
193 625 630 635 640
195 atg cca ggg gga cgc tgg tgc cct gga gcc aac atc tgc ttg ccg ctg 1968
196 Met Pro Gly Gly Arg Trp Cys Pro Gly Ala Asn Ile Cys Leu Pro Leu
197 645 650 655
198 gac gcc tcc tgc cac ccc cag gcc tgc gcc aat ggc ttc aag tca ggg 2016
199 Asp Ala Ser Cys His Pro Gln Ala Cys Ala Asn Gly Cys Thr Ser Gly
200 660 665 670
202 cca gga cta ccc ggg gcc ccc tat ggc cta tgg aga gag ttc cta ttc 2064
203 Pro Gly Leu Pro Gly Ala Pro Tyr Ala Leu Trp Arg Gln Phe Leu Phe
204 675 680 685
206 tcc gtt ccc ggc ggg ccc ccc ggc cac tac tgg gtc aac ctc cac ggc 2112
207 Ser Val Pro Ala Gly Pro Pro Ala Gln Tyr Ser Val Thr Leu His Gly
208 690 695 700
210 cag gat gtc ctc ata ctc cct ggt gac ctc gtt ggc tgg cag cac gac 2160
211 Gln Asp Val Leu Met Leu Pro Gly Asp Leu Val Gly Leu Gln His Asp
212 705 710 715 720
214 gct gga cct ggc gcc ctc cta cac tgc tgg ccc gct ccc ggc cac cct 2208
215 Ala Gly Pro Gly Ala Leu Leu His Cys Ser Pro Ala Pro Gly His Pro
216 725 730 735
218 ggt cca gga gcc ccc tac ctc tcc gcc aac gcc tgg tca tgg ctg ccc 2256
219 Gly Pro Arg Ala Pro Tyr Leu Ser Ala Asn Ala Ser Ser Trp Leu Pro
220 740 745 750
222 cac atg cca gcc cac ctg gag gcc aat tgg ggc tgc cct gcc tgt gcc 2304
223 His Leu Pro Ala Gln Leu Gln Gly Thr Trp Gly Cys Pro Ala Cys Ala
224 755 760 765
226 ctg cga ctg ctt gca caa cgg gaa cag ctc acc gtg ctg ctg ggc ttg 2352
227 Leu Arg Leu Leu Ala Gln Arg Gln Gln Leu Thr Val Leu Leu Gly Leu
228 770 775 780
230 agg cca aac cct gga ctg cgg ctg cct ggg cgc tat gag gtc cgg gca 2400
231 Arg Pro Asn Pro Gly Leu Arg Leu Pro Gly Arg Tyr Gln Val Arg Ala
232 785 790 795 800
234 gag gta gcc aat ggc gtg tcc agg cac aac ctc tcc tcc agc ttt gac 2448
235 Glu Val Gly Asn Gly Val Ser Arg His Asn Leu Ser Cys Ser Phe Asp
236 805 810 815
238 gtg gta tcc cca gtg gct ggg ctg cgg gtc atc tac cct gcc ccc cgc 2496
239 Val Val Ser Pro Val Ala Gly Leu Arg Val Ile Tyr Pro Ala Pro Arg
240 820 825 830
242 gac gga cgc ctc tac gtg ccc acc aac ggc tca gcc tgg gtg ctc cag 2544
243 Asp Gly Arg Leu Tyr Val Pro Thr Asn Gly Ser Ala Leu Val Leu Gln
244 835 840 845
246 gtg gac tct ggt gcc aac gcc aag gcc aag gct cgc tgg cct ggg ggc 2592
247 Val Asp Ser Gly Ala Asn Ala Thr Ala Thr Ala Arg Trp Pro Gly Gly
248 850 855 860
250 agt ctc agc gcc cgc ttt gag aat gtc tgc cct gcc ctg gtg gcc acc 2640
251 Ser Leu Ser Ala Arg Phe Glu Asn Val Cys Pro Ala Leu Val Ala Thr
252 865 870 875 880
254 ttc gtg ccc gcc tgc ccc tgg gag acc aac gat acc ctg ttc tca gtg 2688
255 Phe Val Pro Ala Cys Pro Trp Glu Thr Asn Asp Thr Leu Phe Ser Val

```

RAW SEQUENCE LISTING

DATE: 01/16/2003

PATENT APPLICATION: US/09/655,160

TIME: 13:57:46

Input Set : A:\PTO.VSK.txt

Output Set: N:\CRF4\01162003\I655160.raw

857	885	890	895	
859 gta gca ctg cgg tgg ctc agt gag ggg gag cac gtg gtg cac gtg gtg	2736			
860 Val Ala Leu Pro Trp Leu Ser Glu Gly Glu His Val Val Asp Val Val				
861	900	901	910	
861 gtg gaa aac agc ggc agc cgg gcc aac ctc agc ctg cgg gtg aag gcc	2744			
862 Val Glu Asn Ser Ala Ser Arg Ala Asn Leu Ser Leu Arg Val Thr Ala				
863	915	920	925	
863 gag gag ccc atc tat ggc ctc cgg gcc aag ccc agc gcc gag gcc cgt	2836			
864 Glu Glu Pro Ile Cys Gly Leu Arg Ala Thr Pro Ser Pro Glu Ala Arg				
864	930	935	940	
871 gta ctg ccc gaa ggc cta gtg agc tac agc ccc ctg gtg gag gcc gcc	2840			
871 Val Leu Glu Gly Val Leu Val Arg Tyr Ser Pro Val Val Glu Ala Gly				
872	945	950	955	960
872 cgg gac atg gtc ttc cgg tgg acc atc aac gac aag gag tcc ctg acc	2848			
873 Ser Asp Met Val Phe Arg Trp Thr Ile Asn Asp Lys Gln Ser Leu Thr				
873	965	970	975	
873 ttc cag aac gtg gtc ttc aat gtc att tat cag agc gcc gcc gtc ttc	2856			
874 Phe Gln Asn Val Val Phe Asn Val Ile Tyr Gln Ser Ala Ala Val Phe				
874	980	985	990	
883 aag ctc tca ctg aag gcc tcc aac cac gtg agc aac gtc acc gtg aac	2864			
883 Lys Leu Ser Leu Thr Ala Ser Asn His Val Ser Asn Val Thr Val Asn				
884	995	1000	1005	
887 tac aac gta acc ctg gag cgg atg aac agc atg cag ggt ctg cag	2868			
887 Tyr Asn Val Thr Val Glu Arg Met Asn Arg Met Gln Gly Leu Gln				
888	1010	1015	1020	
891 gtc tcc aca ctg cag gcc gtc ctg tcc ccc aat gcc aag cta gca	2874			
891 Val Ser Thr Val Pro Ala Val Leu Ser Pro Asn Ala Thr Leu Ala				
892	1025	1030	1035	
893 ctg aag gcc ggc ctg ctg gtc gac tgg gcc gtg gag gtg gcc ttc	2880			
893 Leu Thr Ala Gly Val Leu Val Asp Ser Ala Val Glu Val Ala Phe				
894	1040	1045	1050	
899 ctg tgg acc ttc tgg gat gga gag cag gcc ctc ccc cag ttc cag	2884			
899 Leu Trp Thr Phe Gly Asp Gly Glu Gln Ala Leu His Gln Phe Gln				
900	1055	1060	1065	
901 cct cgg tac aac cag tcc ttc cca gtt cca gac ccc tgg gtg gcc	2888			
901 Pro Pro Tyr Asn Glu Ser Pro Pro Val Pro Asp Pro Ser Val Ala				
902	1070	1075	1080	
907 cag gtg ctg gtg gag cac aat gtc aag ccc acc tac gct gcc cca	2894			
907 Gln Val Leu Val Glu His Asn Val Thr His Thr Tyr Ala Ala Pro				
908	1085	1090	1095	
911 ggt gag tac ctc ctg acc gtc ctg gca tct aat gcc ttc gag aac	2898			
911 Gly Glu Tyr Leu Leu Thr Val Leu Ala Ser Asn Ala Phe Glu Asn				
912	1100	1105	1110	
915 ctg aag cag cag ctg cct gtc agc gtg cgc gcc tcc ctg ccc tcc	2904			
915 Leu Thr Gln Gln Val Pro Val Ser Val Arg Ala Ser Leu Pro Ser				
916	1115	1120	1125	
919 gtg gct gtg ggt gtg agt gac gcc ctc ctg gtg gcc ggc cgg ccc	2908			
919 Val Ala Val Gly Val Ser Asp Gly Val Leu Val Ala Gly Arg Pro				
920	1130	1135	1140	

RAW SEQUENCE LISTING ERROR SUMMARY

DATE: 01/16/2003

PATENT APPLICATION: US/09/655,160

TIME: 13:57:48

Input Set : A:\PTO.VSK.txt

Output Set : N:\CRF4\01162003\I655160.raw

Invalid Line Length:

The rules require that a line not exceed 72 characters in length. This includes spaces.

Seq#:1; Line(s) 6

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/655,160

DATE: 01/16/2003

TIME: 13:57:48

Input Set : A:\PTO.VSK.txt

Output Set: N:\CRF4\01162003\I655160.raw

L:12 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:34 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:1,Line#:32